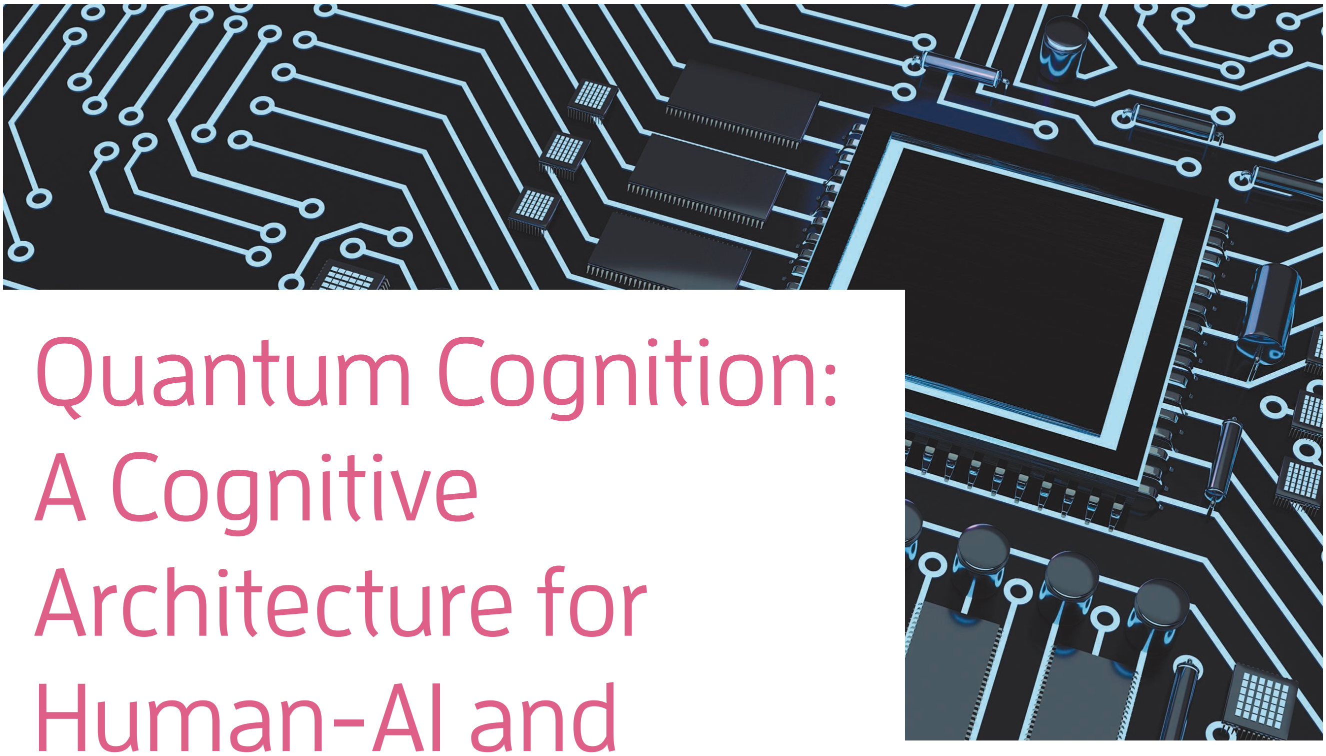
EDITOR JOSEPH WILLIAMS
Infosys Consulting;
joseph.williams@infosys.com



# Quantum Cognition: A Cognitive Architecture for Human-AI and In-Memory Computing

**Fariborz Farahmand**, Georgia Institute of Technology

*This article focuses on human-artificial intelligence (AI): "machines that think, that learn and that create". I shed light on some issues that have led to unbalanced progress in AI (more progress in artificial and less progress in intelligence), and introduce quantum cognition as a viable cognitive architecture for human-AI and emerging hardware.*

*Digital Object Identifier 10.1109/MC.2023.3242056*
*Date of current version: 5 April 2023*

Most of the cognitive architectures, that is, models of human reasoning in artificial intelligence (AI) research, do not necessarily try to model the human reasoning process. They assume humans are rational agents, that is, utility maximizers, who *always* follow Boolean logic, which implies that events can *always* be combined (for example, via logical conjunction) in any order. They try to locate an equivalent of an arithmetic logic unit (ALU) in human brains, and shuffle data to make it independent and identically distributed (IID). The following section shows more details.

## ISSUES WITH CLASSICAL PROBABILITY IN REASONING

Classical probability theory and widely accepted Kolmogorov axioms follow Boolean logic. This implies that the logic of events is commutative and that events are always *compatible*. That is, A and B is the same as B and A, and simultaneous measurements of A and B, or B and A will cause no interference.

In fact, this logic works well for compatible events. For example, first measuring your height and then your weight, or first measuring your weight and then your height, or simultaneously measuring your height and weight all yield the same result.

But the reality is that events could be *incompatible*, that is, evaluation is order dependent, and interference could occur. For example, consider question A: "Are you going to Florida?" and question B: "Did you hear there is a storm coming to Florida?" First requesting an answer to question A and then question B, or first requesting an answer to question B and then question A, or asking questions A and B simultaneously and then requesting an answer could result in different answers.

Additionally, a Boolean conjunction can be perceived as more representative than one of its constituents and change human reasoning. Here is a simplified example.

Assume that Bob has been identified as the suspect of exploiting a zero-day vulnerability. Also, assume that exploiting such a vulnerability has often been observed from the members of a famous hacking group called H. Then, which of the following scenarios seems more probable?

1. Bob is a skilled hacker.
2. Bob is a skilled hacker, and a member of group H.

Intuitively speaking, scenario 2 could be perceived as more probable. But with Boolean logic and classical probability theory, the probability of two events happening together cannot be larger than the probability of a single event. We perceive scenario 2 as being more probable than scenario 1 because of a conjunction fallacy, a cognitive bias identified by Tversky and Kahneman[1] that explains that humans are usually more inclined to believe a detailed story with explicit details over a short compact one. In fact, phishing attackers have greatly benefited from this bias by first providing their targets with an explicit and detailed description of an event that requires immediate attention and then asking them to click a link.

## UNCONSCIOUS LEARNING

Goyal and Bengio[2] argue that to achieve human-AI, we need to move from system 1/implicit/unconscious processing to system 2/explicit/conscious processing. System 1 operation is similar to when we are driving in a familiar neighborhood, where we can be fast and unconscious. System 2 operation is similar to when we are driving in an unfamiliar neighborhood and need to be slow and conscious and may need consultation as well. Goyal and Bengio's[2] proposal requires "sequential conscious processing" and considering "attention as sequentially selecting what computation to perform on what quantities." However, as briefly discussed, classical probability has major limitations with sequential processing. It assumes that all events are compatible and does not consider order effect.

For example, to avoid overfitting (paying too much attention to the particular dataset it is trained on), the machine learning community shuffles data to make them IID. But the reality is that data do not arrive to us as IID.[2]

> "Nature doesn't shuffle data, and we should not. When we shuffle the data, we destroy useful information about those changes in distribution that are inherent in the data we collect and contain information about causal structure."

## QUANTUM PROBABILITY FOR REASONING AND INFERENCE

I recommend quantum cognition[3] as a viable alternative for cognitive architectures that use classical reasoning and inference. Quantum cognition is different from the quantum mind. It does not follow the assumption that there is something quantum like taking place in the brain but takes inspiration from the mathematical structure of quantum theory and its dynamic principles. For example, it uses quantum probability—modeling cognition using the theory of probability from quantum mechanics, without any of the physics.

The following section shows an example feature of quantum probability that makes it appropriate for human-AI software and hardware.

## CAPTURING INCOMPATIBILITIES

Quantum probability, unlike classical probability that assumes all questions are compatible, can capture incompatibles. Quantum probability uses vector space and subspace similar to classical probability's use of sample space and event (that is, a subset of sample space), respectively. Vector space contains all possible outcomes for questions. A vector representing a question outcome spans a 1D subspace, called a *ray*, and the set of beliefs a person has about the question is represented by a unit length vector, called a *state vector*. Quantum probability also uses a mapping process, called *projecting*, and the probability assigned to an event equals the squared length of the projection. To compute the conjunction of question outcomes, quantum probability employs a sequential projection. This allows distinguishing between orders, that is, project A and then project B has a different outcome than project B and then project A.

### Revisiting the simplified example

Here, we revisit our simplified example to illustrate how quantum probability, using vector space, can illustrate conjunction fallacy in human reasoning. In Figure 1, blue arrows represent "Bob being a skilled hacker" by ***B*** and its negation with ***≁B***. Similarly, orange arrows represent "being a member of group H" with ***H*** and its negation with ***≁H***. ***S***, the state vector, represents our belief state about Bob's characterization and is represented by the black arrow. In Figure 1, projection paths are shown by green and red dotted lines. Probabilities are computed as the squared length of the projection of the state vector onto the corresponding axis and shown by green and red square lengths. The projection onto the B ray is shown by the green dotted line, and the probability of (***B***) equals the squared length of this bar, shown by the green square length. For the probability of (***B*** and ***H***), we need to follow two steps, as shown by the two red dotted lines. First, we project the state vector onto the H ray. Second, we project this previous projection onto the B ray. Then, the probability of (***B*** and ***H***) is the squared length of the last projection, shown by the red square length.

In Figure 1, the sequential probability of (***B*** and ***H***) is greater than the probability of a single event, that is, the probability of (***B***), corresponding to the red square length being longer than the green square length. This is because of the conjunction fallacy that led to perceiving scenario 2 as being more probable than scenario 1. We can relate the incompatibility of (***B*** and ***H***), leading to their interference, to the representativeness heuristic (a mental shortcut); conjunction seems more representative than one of its constituents, and being a member of H can be easier to imagine or to retrieve than Bob as an inclusive category. For a mathematical explanation of this example, see the supplementary materials available at 10.1109/MC.2023.3242056.

The ability of quantum probability to capture incompatibilities can also play an important role in developing causal structures for human-AI, specifically when we are dealing with incompatible events by putting together complex situations with massive amounts of data from various sources. In such situations, we need causal structural models to uncover the underlying mechanisms of the data versus elemental causal induction, that is, modeling a single cause-and-effect relationship, using classical probability. In such complex situations, quantum probability can provide a way to formalize the idea of structurally local causal reasoning by working with incompatible events, pasting together sample spaces, and forming a vector space.

For example, assume we need to make a predictive judgment, that is, find the conditional probability of an effect given a cause, or P (effect|cause), in a complex problem with massive amounts of data, where the order of data arrival matters. Quantum probability enables us to break the problem into smaller problems by answering queries such as:

P (effect|cause$_1$, no alternative cause),
P (effect|cause$_1$, cause$_2$),
P (effect|cause$_2$, cause$_1$), etc.

I recommend quantum cognition as a viable alternative for cognitive architectures that use classical reasoning and inference.

## QUANTUM PROBABILITY FOR IN-MEMORY COMPUTING

Quantum probability uses vector space, similar to the computing framework vector symbolic architectures (VSAs), also known as *hyperdimensional computing*, which is central to the emerging hardware, for example, in-memory computing (IMC). In conventional von Neumann architecture, memory and processor are separate, and the computation requires data to be moved back and forth. But with IMC architecture using "vector-matrix multiplication,"[4] memory and processor are fused together, and computations are performed where data are stored with minimal data movement. That makes IMC, in contrast to conventional von Neumann architecture, similar to the human brain, where memory and computation are collocated. In fact, locating the equivalent of an ALU in the human brain is an unrealistic expectation.

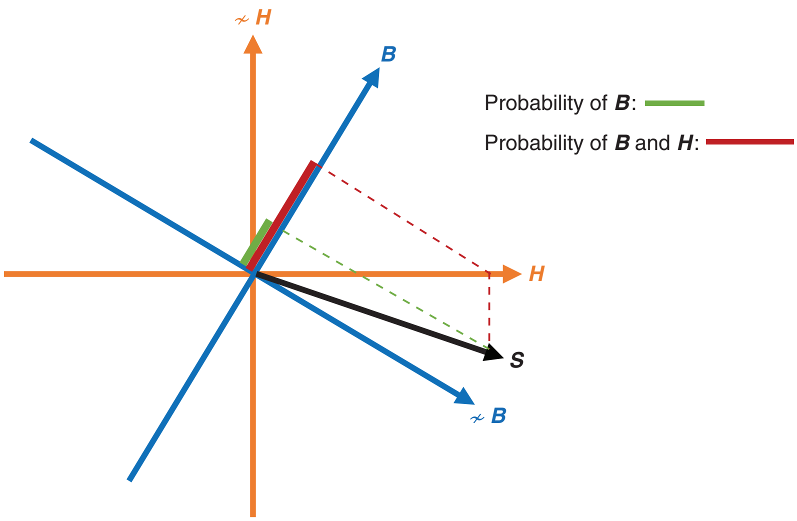



FIGURE 1. A quantum probability representation of the simplified example ***B***: "Bob being a skilled hacker" and ***H***: "being a member of group H". ***S*** represents the state vector, our belief state about Bob's story.

Quantum probability and IMC can be considered as promising computational architectures for human-AI as both use VSAs. So it is reasonable to consider quantum probability as a cognitive architecture for IMC.

Here is an example. Working memory in the human brain is a mechanism for the temporary storage of information related to the current task. It is critical for cognitive capacities such as attention, reasoning, and learning; thus, most cognitive architectures implement it in some form. With quantum cognition, we can use high-dimensional vectors to represent the function of working memory and to deal with the relevant data in an ongoing computation. Quantum probability's state vector can be considered as a working-memory state that represents human beliefs about feature patterns and serves as a cache for the current world model, the state of the system, and/or current goals.

Quantum probability builds a strong mathematical foundation for IMC and organizing the "operations on hyperdimensional patterns that could be used for computing." By viewing patterns as vectors, we can[5]

> "tap into the vast body of knowledge about vectors, matrices, linear algebra, and beyond. This indeed has been the tradition in artificial neural-net research, yet rich areas of high-dimensional representation remain to be explored."

Agents who *always* maximize utility, using structures that *always* follow Boolean logic, are fundamental to existing AI computing architectures. But I argue that we need to learn from Einstein's relativity theory, Gödel's incompleteness theorem, and Simon's bounded rationality theory, as they all shed light on the collapse of absolutes.

In this article, I presented some computational limitations of existing AI systems. I explained that, unlike the axioms of classical probability, the logic of events is not necessarily Boolean. If two events A and B are incompatible, then the conjunction of events A and B cannot be defined because they do not commute, in sharp contrast with Boolean logic, where events always commute.

Quantum probability and IMC can be considered as promising computational architectures for human-AI as both use VSAs.

I offered recommendations about quantum probability and explained how to consider quantum states as measures over the non-Boolean structure of projection operators. To compare quantum states and classical probabilistic states, I explained how projection can be used to describe experiments similar to classical probability. I explained how causal structural models (versus elemental causal induction) can help with capturing sequential conscious processing. I also explained how quantum probability's use of vector space makes it an appropriate cognitive architecture for IMC architecture.

Achieving human-AI and developing "machines that think that learn and that create"[6] require computational models that can act likewise. But human thinking, learning, and creating are often strongly context and order dependent, and that appears perplexing to classical probability and utility maximization models.


## ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under Award 2041788. A brief mathematical explanation of the simplified exam and its revisit can be found in the supplementary materials available at 10.1109/MC.2023.3242056.


## REFERENCES


1. A. Tversky and D. Kahneman, "Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment," *Psychol. Rev.*, vol. 90, no. 4, pp. 293–315, Oct. 1983, doi: 10.1037/0033-295X.90.4.293.
2. A. Goyal and Y. Bengio, "Inductive biases for deep learning of higher-level cognition," *Proc. Roy. Soc. A*, vol. 478, no. 2266, Oct. 2022, Art. no. 20210068, doi: 10.1098/rspa.2021.0068.
3. J. R. Busemeyer and P. Bruza, *Quantum Models of Cognition and Decision*. Cambridge, U.K.: Cambridge Univ. Press, 2014.
4. S. Spetalnick and A. Raychowdhury, "A practical design-space analysis of compute-in-memory with SRAM," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 69, no. 4, pp. 1466–1479, Apr. 2022, doi: 10.1109/TCSI.2021.3138057.
5. P. Kanerva, "Hyperdimensional computing: An introduction to computing in distributed representation with high-dimensional random vectors," *Cogn. Comput.*, vol. 1, pp. 139–159, Jun. 2009, doi: 10.1007/s12559-009-9009-8.
6. H. Simon and A. Newell, "Heuristic problem solving: The next advance in operations research," *Oper. Res.*, vol. 6, no. 1, pp. 1–10, Jan./Feb. 1958. [Online]. Available: https://www.jstor.org/stable/167397



**FARIBORZ FARAHMAND** is a research faculty member at the School of Electrical and Computer Engineering, Georgia Institute of Technology (Georgia Tech), Atlanta, GA 30332 USA. He is a Senior Member of IEEE. Contact him at fariborz@ece.gatech.edu.